# 小立碗藓 LysM 型类受体激酶基因家族生物信息学分析

高梅[1]，辛健康[1]，姜山[2]*

（1. 贵州师范大学 生命科学学院，贵阳 550001；2. 贵州师范大学 国际教育学院，贵阳 550001）

**摘要：**植物 LysM 型类受体激酶（lysin motif receptor-like kinase，LYKs）是植物中发现的一类重要的 RLK，在植物生长发育、抵御逆境胁迫等方面具有不可忽视的作用，是植物中基因功能的研究热点。为更好地了解小立碗藓中的 *LYK* 基因，该文利用生物信息学的方法对小立碗藓（*Physcomitrella patens*）LysM 型类受体激酶基因家族成员进行鉴定及分析。通过分析小立碗藓 LYK 家族成员的基本物理信息、基因结构、染色体定位及系统发生关系，初步探讨了其 *LYK* 基因结构、进化与功能间的联系。结果表明：小立碗藓中共有 21 个 *LYK* 基因，其氨基酸序列大小在 625~755 aa 之间，分子量范围为 69.54~82.02 kDa，等电点在 5.98~7.78 之间。将小立碗藓所有 LYK 蛋白与 3 种典型模式植物（水稻、拟南芥和蒺藜苜蓿）的 LYK 蛋白共同构建系统进化树，所有 LYK 蛋白被分为 4 个亚组（LYK-I、LYK-II、LYR-I 和 LYR- II）。小立碗藓各亚组内成员的基因结构、保守域特征显示出较为相似的特征，由此推测其可能具有相同或相似的功能。染色体定位发现，21 个 *LYK* 基因集中分布于 4 条染色体上并出现小型基因簇，这可能同样与基因功能相联系。该文分析了小立碗藓 LysM 型类受体蛋白激酶基因家族的基本信息，可为后续深入研究其 *LYK* 基因家族成员的生理生化功能奠定基础。

**关键词：**小立碗藓，LysM 型类受体激酶，基因家族，生物信息学，基因结构，染色体定位，系统进化关系

**中图分类号：**Q786　　　　**文献标志码：**A

# Bioinformatics analysis of lysin motif receptor-like kinase gene family in *Physcomitrella patens*

GAO Mei[1], XIN Jiankang[1], JIANG Shan[2]*

(1.*School of Life Science* , *Guizhou Normal University*, Guiyang, 550001, China; 2. *School of International Education, Guizhou Normal University*, Guiyang, 550001, China )

**Abstract:** Plant lysin motif receptor-like kinase (LYKs) are an important type of RLK found in plants. They have a non-negligible role in plant growth and development, resistance to adversity stress, etc. It is a gene function in plants research hotspots. In order to better understand the LYK gene in *Physcomitrella patens*, this article uses bioinformatics to identify and analyze members of the LysM-type receptor kinase gene family of *Physcomitrella patens*(*P. patens*). By analyzing the basic physical information, gene structure, chromosomal location and phylogenetic relationship of the LYK family members, the relationship between LYK gene structure, evolution and function was preliminarily discussed. The results show that there are 21 LYK genes in the moss, the amino acid sequence size ranges from 625 to 755aa, the molecular weight ranges from 69.54 to 82.02kDa, and the isoelectric point ranges from 5.98 to 7.78. A phylogenetic tree was constructed by combining all the LYK proteins of *Physcomitrella patens* with the LYK proteins of three typical model plants (*Rice*, *Arabidopsis*, and *Medicago truncatula*). All LYK proteins were divided into 4 subgroups (LYK-I, LYK- II, LYR-I and LYR-II). The genetic

---

**收稿日期：**2020-12-09

**基金项目：**国家自然科学基金项目(30860158；31560508；31260426)；[Supported by the National Natural Science Foundation of China (31560508;31260426;30860158)]。

**作者简介：**高梅（1995-），女，山东烟台人，硕士研究生，主要从事植物分子生物学研究，（Email）1024621800@qq.com。

***通信作者：**姜山，博士，教授，现主要从事植物病理学研究，（Email）kyosan200312@hotmail.com。

structure and conserved domain characteristics of the members in each subgroup of *Physcomitrella patens* showed relatively similar characteristics, and it was speculated that they might have the same or similar functions. Chromosome mapping found that 21 LYK genes were concentrated on 4 chromosomes and there were small gene clusters, which may also be related to gene function. This article analyzes the basic information of the LysM receptor-like protein kinase gene family of *Physcomitrella patens*, which can lay the foundation for the in-depth study of the physiological and biochemical functions of its LYK gene family members.

**Key words:** *Physcomitrella patens*, lysin motif receptor-like kinase, gene family, bioinformatics, gene structure, chromosome location, phylogenetic relationship

植物类受体蛋白激酶是植物中一类重要的蛋白激酶，这些蛋白激酶与动物中受体蛋白激酶的分子结构类似，由于与其特异结合的胞外信号分子或配体尚未被鉴定，因此被称为类受体蛋白激酶（Receptor-like protein kinase，RLK）。典型的 RLK 由胞外结构域、跨膜结构域和胞内激酶域三部分构成，其中胞外结构域含特异性结构基元，能够特异性识别细胞信号分子（田丽梅，2016；Lehti-Shiu et al., 2009）。其中植物 LysM 型类受体激酶（Lysin Motif Receptor-like Kinase，LYKs）是植物中发现的一类重要的类受体蛋白激酶。LysM 型类受体蛋白激酶（LysM Receptor-like Kinase，LYKs）的典型特征是其胞外结构域中含有 1-3 个 LysM 结构基元（Afzal et al., 2008）。

不同植物中 LYKs 家族成员数存在差异，拟南芥和水稻中分别含有 5 和 10 个 LysM-RLKs（Zhang , 2007；Shiu et al., 2004），同时不同植物中 LYKs 胞外含有的 LysM 结构基元数也有所不同。除古细菌外，大多数生物中 LysM 蛋白结构域长度约为 40aa。植物中 LYKs 家族成员根据其激酶结构域的不同，可以划分为两种主要类型：LYK 或 LysM-I 和 LYR 或 LysM-II，其中 LYK 亚家族又可分为 LYK-I 和 LYK-II；LYR 亚家族分为 LYR-I 和 LYR-II（Limpens et al., 2003; Arrighi et al., 2006）。目前，大多数 LysM-RLK 能够感知病原菌分子或参与植物-微生物间的相互作用，包括建立防御反应或根内共生体（Wan et al., 2008; Radutoiu et al., 2003）。

植物 LysM 型类受体蛋白激酶在植物防御反应及与微生物共生过程中的重要作用已被证实，但少有对小立碗藓中 LysM-RLK 家族基因功能的研究报道。小立碗藓(*Physcomitrella patens*)是非维管束植物中的模式植物，也是早期登陆陆生植物代表，小立碗藓基因组大小为 511 Mb，共含 27 条染色体且全基因组测序已完成(Rensing et al., 2002)，基于此，本文采用生物信息学手段，在基因组范围鉴定小立碗藓 LYKs 基因家族成员，并对其进行蛋白序列分析、结构域预测、染色体定位及进化分析等，意在为后续研究该基因家族成员的生理生化功能提供参考。

# 1 研究对象与方法

## 1.1 研究对象及序列数据下载

本研究以模式植物小立碗藓为研究对象，小立碗藓基因组注释序列下载于 Ensembl Plants 数据库（http://plants.ensembl.org/index.html）（Kersey, 2018），拟南芥基因组注释序列从 TAIR 数据库（https://www.arabidopsis.org/）下载（Poole, 2007）。水稻基因组注释序列下载于国家水稻数据中心（http://www.ricedata.cn/gene/）（樊颖伦，2009），蒺藜苜蓿基因组注释序列从 Phytozome v12.1 数据库（https://phytozome.jgi.doe.gov/pz/portal.html）下载。

## 1.2 小立碗藓 *LYK* 基因家族成员的鉴定

利用 TAIR 数据库中拟南芥 *LYK* 基因家族成员（共 5 个）蛋白序列作为参考序列集合，在小立碗藓基因组序列（http://asia.ensembl.org/index.html）中进行同源搜索，查找小立碗藓中 *LYK* 基因家族成员，并进一步利用 NCBI Conserved Domain Search (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) 筛选，确保氨基酸序列同时存在 N-末端 LysM 和 C-末端 Pkinase 结构域，最终得到 21 条候选 *LYK* 基因，用 TBtools 软件对基因保守结构域可视化（Chen et al., 2018）。

## 1.3 小立碗藓 *LYK* 基因家族的蛋白特征分析及亚细胞定位预测

将鉴定得到的 21 条候选类受体蛋白激酶家族蛋白序列，利用 ExPASy 网站（http://expasy.org/）进行分子量、理论等电点的预测，并利用 SMART 网站(http://smart.embl-heidelberg.de/)对所有基因进行保守结构域

位置预测（陈紫华，2019）；利用 CELLO 在线网站（http://cello.life.nctu.edu.tw/）预测小立碗藓 *LYK* 基因家族的亚细胞定位，图中*表示可信度最高（宋希茜，2016）。

### 1.4 小立碗藓 LYK 蛋白的保守域结构分析

利用 MEME 网站（http://meme-suite.org/）基于相似度较高的基序，即 Motif，分析 *LYK* 基因家族成员蛋白序列保守特征，用 TBtools 软件进行可视化。

### 1.5 小立碗藓 *LYK* 基因家族成员基因结构分析和基因组定位分析

根据检索到的小立碗藓 *LYK* 基因家族成员的基因组信息，利用从 EnsemblPlants 数据库(http://plants.ensembl.org/index.html)下载的小立碗藓基因染色体定位文件（.gff3），获得小立碗藓 *LYK* 基因家族成员的外显子及内含子分布情况和它们在小立碗藓基因组染色体上的位置信息，最后利用 TBtools 软件绘制基因结构图和染色体定位图。

### 1.6 小立碗藓 *LYK* 基因家族的进化分析

对小立碗藓 *LYK* 基因家族成员构建系统进化树：利用进化树分析软件 MEGA7.0 中的 Muscle 对小立碗藓 LYK 蛋白序列进行多序列比对（Edgar，2004）。利用进化树分析软件 MEGA7.0，执行参数 Partial deletion，选用合适的氨基酸替代模型（LG+G+I），使用基于 LG 矩阵模型的最大似然法推断进化历史（Sagulenko et al., 2018）。执行参数 Poission correction、partial deletion，校验参数 bootstrap= 1 000，构建系统进化树，对小立碗藓 *LYK* 基因进行初步的亚族分类。

对小立碗藓 *LYK* 基因家族成员与水稻、拟南芥和蒺藜苜蓿中的同源家族成员构建系统进化树，方法同上，对小立碗藓 *LYK* 基因进一步分类。

## 2 结果与分析

### 2.1 小立碗藓 *LYK* 基因家族成员基本信息

利用 TAIR（https://www.arabidopsis.org/）数据库下载的拟南芥 *LYK*（LysM-RLK）家族基因序列作为基因家族的参考序列（共 5 个），使用 TBtools 软件中的 Blast Wrapper，调取小立碗藓中的可能的 LYKs 基因家族成员序列，得到 160 个候选基因，再利用 NCBI BLASTp 进行双向 Blast，去重复得到 50 条候选 *LYK* 基因，最后利用 NCBI Conserved Domain Search（https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi）基于保守结构域进一步筛选，最终得到 21 条候选 *LYK* 基因，TBtools 软件可视化如图 1。利用生物信息学方法，对小立碗藓 *LYK* 基因家族成员的蛋白长度、染色体位置、分子量、理论等电点等基本信息进行分析并编号，详见表 1。

通过蛋白特征分析，小立碗藓 LYK 蛋白的分子量为 69.54~82.02 kDa，变化较稳定，理论等电点最小为 5.98，最大为 7.78。这些 *LYK* 基因编码蛋白的氨基酸长度从 625 ~755 aa 之间波动。运用 SMART 网站鉴定和 Pfam 数据库(http://pfam.xfam.org)验证结果显示，21 个 LysM 型类受体蛋白激酶同时具有 PKc_like 和 LysM 特征结构域，不同 LYK 所含 LysM domain 的数目存在差异。这一结果与拟南芥中已鉴定出的 LYK 蛋白特征相似（Shiu et al., 2004）。亚细胞定位显示，通过 CELLO 网站预测小立碗藓 *LYK* 基因家族成员全部位于质膜。

表 1 小立碗藓 LYK 蛋白家族成员基本信息
Table.1 Identification of LYK protein gene family members from *P.patens*

| 基因编号 Gene name | 基因号 Gene ID | 蛋白长度 Length of protein（aa） | 理论等电点 Theoretical isoelectric point（Pi） | 内含子数目 Intron | 分子量（kDa） Molecular weight（Mw） | 亚细胞定位 Subcellular localization CELLO 预测 CELLO prediction |
|---|---|---|---|---|---|---|
| *PpLYK1* | Pp3c7_26350V3.1 | 678 | 6.65 | 1 | 74.13 | 质膜 plasma membrane（4.002 *） |
| *PpLYK2* | Pp3c7_26350V3.9 | 716 | 7.78 | 3 | 78.11 | 质膜 plasma membrane（4.101 *） |
| *PpLYK3* | Pp3c7_26350V3.7 | 678 | 6.65 | 2 | 74.13 | 质膜 plasma membrane（4.002 *） |
| *PpLYK4* | Pp3c7_26350V3.8 | 678 | 6.65 | 1 | 74.13 | 质膜 plasma membrane (4.002 *) |
| *PpLYK5* | Pp3c7_26350V3.10 | 716 | 7.78 | 2 | 78.11 | 质膜 plasma membrane (4.101 *) |
| *PpLYK6* | Pp3c1_6050V3.1 | 671 | 6.24 | 1 | 72.56 | 质膜 plasma membrane (3.086 *) |
| *PpLYK7* | Pp3c1_6050V3.2 | 755 | 6.04 | 3 | 82.02 | 质膜 plasma membrane (3.837 *) |
| *PpLYK8* | Pp3c1_6050V3.3 | 755 | 6.04 | 2 | 82.02 | 质膜 plasma membrane (3.837 *) |
| *PpLYK9* | Pp3c11_6330V3.3 | 640 | 7.11 | 13 | 70.45 | 质膜 plasma membrane (2.579 *) |
| *PpLYK10* | Pp3c11_6330V3.5 | 640 | 7.11 | 13 | 70.45 | 质膜 plasma membrane (2.579 *) |
| *PpLYK11* | Pp3c11_6330V3.4 | 640 | 7.11 | 13 | 70.45 | 质膜 plasma membrane (2.579 *) |
| *PpLYK12* | Pp3c11_6330V3.6 | 640 | 7.11 | 13 | 70.45 | 质膜 plasma membrane (2.579 *) |
| *PpLYK13* | Pp3c7_25560V3.3 | 625 | 5.98 | 13 | 69.54 | 质膜 plasma membrane (3.864 *) |
| *PpLYK14* | Pp3c7_25560V3.4 | 625 | 5.98 | 12 | 69.54 | 质膜 plasma membrane (3.864 *) |
| *PpLYK15* | Pp3c7_25560V3.1 | 625 | 5.98 | 13 | 69.54 | 质膜 plasma membrane (3.864 *) |
| *PpLYK16* | Pp3c7_25560V3.2 | 625 | 5.98 | 14 | 69.54 | 质膜 plasma membrane (3.864 *) |
| *PpLYK17* | Pp3c9_5820V3.1 | 697 | 6.23 | 1 | 75.27 | 质膜 plasma membrane (2.659 *) |
| *PpLYK18* | Pp3c9_5820V3.2 | 697 | 6.23 | 1 | 75.27 | 质膜 plasma membrane |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | (2.659 *) |
| *PpLYK19* | Pp3c9_5820V3.3 | 697 | 6.23 | 1 | 75.27 | 质膜 plasma membrane (2.659 *) |
| *PpLYK20* | Pp3c9_5820V3.4 | 697 | 6.23 | 2 | 75.27 | 质膜 plasma membrane (2.659 *) |
| *PpLYK21* | Pp3c9_5820V3.5 | 697 | 6.23 | 2 | 75.27 | 质膜 plasma membrane (2.659 *) |

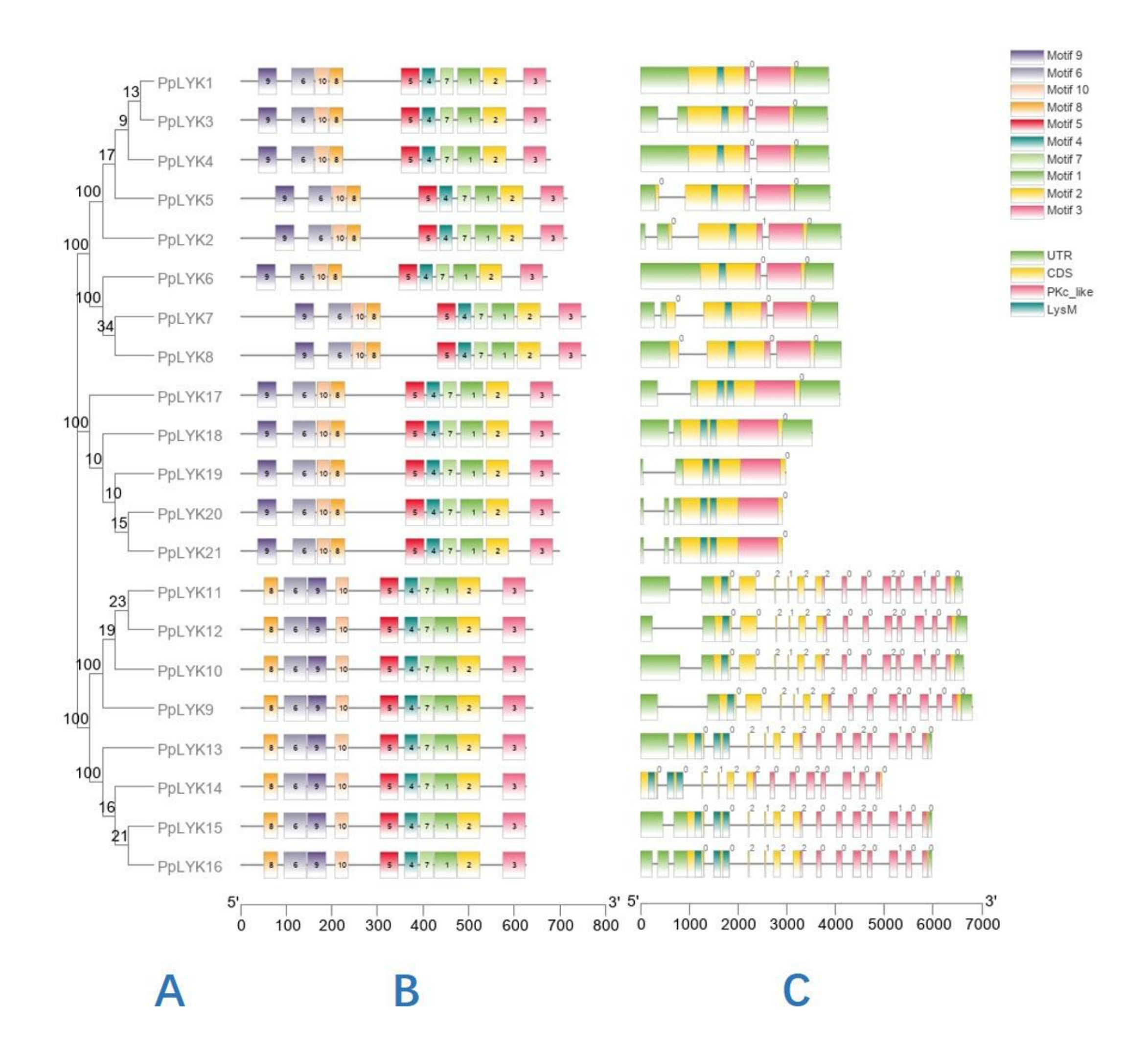
**2.2 小立碗藓 *LYK* 基因结构与蛋白保守结构域分析**

通过 MEME 网站分析小立碗藓 *LYK* 基因序列，得到 10 个保守性较高的 motif（图 1：B），并且 motif 间的排布具有一定规律，主要分为两大类：3ˊ端排布顺序相同，均为 motif 5-motif 4-motif 7-motif 1-motif 2-motif 3 ;5ˊ端排布顺序分为两类：motif 9- motif 6-motif 10-motif 8 和 motif 8-motif 6-motif 9-motif 10。可见，小立碗藓 LYK 家族成员均具有 motif 5-motif 4-motif 7-motif 1-motif 2-motif 3 的相对稳定结构，且位于基因 3ˊ端，但各成员 5ˊ端会出现 motif 排列顺序的差异。MEME 预测的小立碗藓 LYK 蛋白 motif 信息见表 2。发现小立碗藓 LYK 蛋白的 motif 长度最小为 29，最大为 50，并且存在长度相同的 motif。

通过 SMART 网站和 Pfam 网站分析小立碗藓 LYK 蛋白保守结构域，即通常所说的 domain，结果如图 2 和图 1：C。21 个 LYK 蛋白均含有激酶结构域和 LysM 结构域，但所含 LysM 结构域的个数存在一定差异。利用进化树软件 MEGA 7.0 构建小立碗藓 *LYK* 基因家族的系统发育树（图 1：A），发现 21 个 LYK 家族成员从上至下分为 3 个分支，为更准确的分类，后文与已鉴定的 3 种模式植物的 LYK 成员一同构建系统进化树，以更好地了解物种与进化间的联系。

分析 21 个小立碗藓 *LYK* 基因的外显子、内含子结构（图 1：C），发现约 62％的 *LYK* 基因含有 1~3 个内含子；其中有 6 个 *LYK* 基因均含有 13 个内含子，分别为 *PpLYK9*、*PpLYK10*、*PpLYK11*、*PpLYK12*、*PpLYK13*、*PpLYK15*； *PpLYK16* 含有 14 个内含子，为 21 个基因中内含子数量最多的基因，猜测其编码的蛋白生物学功能较为多样。LysM-RLKs 家族中未发现不含内含子的成员，其中 *PpLYK14* 为仅有的不含非编码区的 *LYK* 基因。

将所有 LysM-RLKs 家族成员的保守结构域和基因结构整合在一起（图 1：C），可以发现两类特征结构域（LysM 和 PKc_like）均存在于基因编码区（即 CDS），说明这两类结构域均对基因的功能起作用。图中发现 *PpLYK9* 至 *PpLYK13* 在外显子-内含子结构上与其它 LYK 家族成员存在显著差异。




A. 21 个 LYK 蛋白的系统发育树。用 MEGA7.0 构建系统发育树，使用 21 个 LYK 成员的全长氨基酸序列，选用 Poisson model，Bootstrap=1000。B. *LYK* 基因的保守基序（motif）分布。10 个 motif 用不同颜色矩形框表示，不同 motif 在图中用数字标明。C. *LYK* 基因的保守结构域及外显子、内含子分布。 绿色矩形框表示蛋白的非编码区（即 UTR），黄色矩形框表示蛋白的编码区（即 CDS），中间隔断的黑色线条表示内含子区域，通过底部比例尺可推断各区域长度，数字表示相位。

A. Phylogenetic tree of 21 LYK proteins. Phylogenetic tree was constructed with MEGA7.0, with full-length amino acid sequences of 21 LYK members and poisson model, bootstrap=1000. B. Conserved motifs distribution of LYK genes. 10 motif are represented by rectangular boxes of different colors, and different motif are indicated in figures. C. Conserved domain and exon and intron distribution of *LYK* genes. The green box indicates noncoding region of protein (ie, UTR), the yellow box indicates the coding region of the protein (ie, CDS), and the black line in the middle of the gap indicates the intron region, and the length of each region can be inferred from the bottom scale. Numbers represent phase.

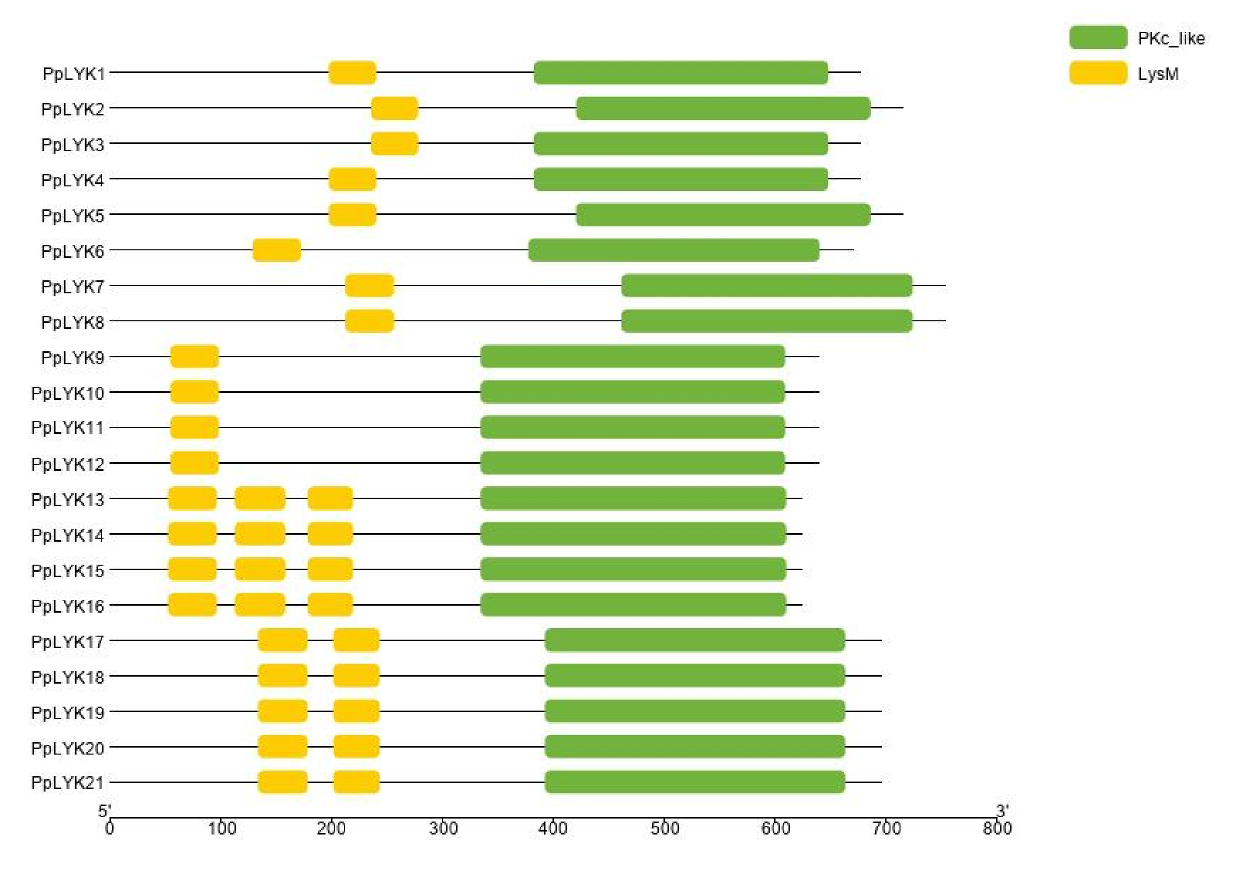
图 1 小立碗藓 LYK 家族成员的系统发育关系、保守基序、保守结构域及基因结构分析

Fig.1 Phylogenetic relationship, conserved motifs， conserved domains and gene structure analysis of the members of the family of LYK in *Physcomitrella patens*

表 2 MEME 预测的 LYK 蛋白 motif 信息

Table.2 MEME predicted LYK proteins motif information

| Motif | E-value | Sites | Width | Logo |
|---|---|---|---|---|
| motif1 | 9.3e-815 | 21 | 50 | |
| motif2 | 1.5e-711 | 21 | 50 | |
| motif3 | 2.4e-590 | 21 | 50 | |
| motif4 | 4.1e-411 | 21 | 29 | |
| motif5 | 1.4e-465 | 21 | 41 | |
| motif6 | 5.3e-528 | 21 | 50 | |
| motif7 | 9.1e-363 | 21 | 29 | |
| motif8 | 1.5e-283 | 21 | 31 | |
| motif9 | 3.6e-361 | 21 | 41 | |
| motif10 | 5.0e-261 | 21 | 29 | |




不同颜色方框表示不同的 domain，各蛋白长度可通过底部比例尺推断。

Different color boxes indicate different domains, and the length of each protein can be inferred by the bottom scale.

图 2 小立碗藓 LYK 蛋白保守结构域（domain）分析

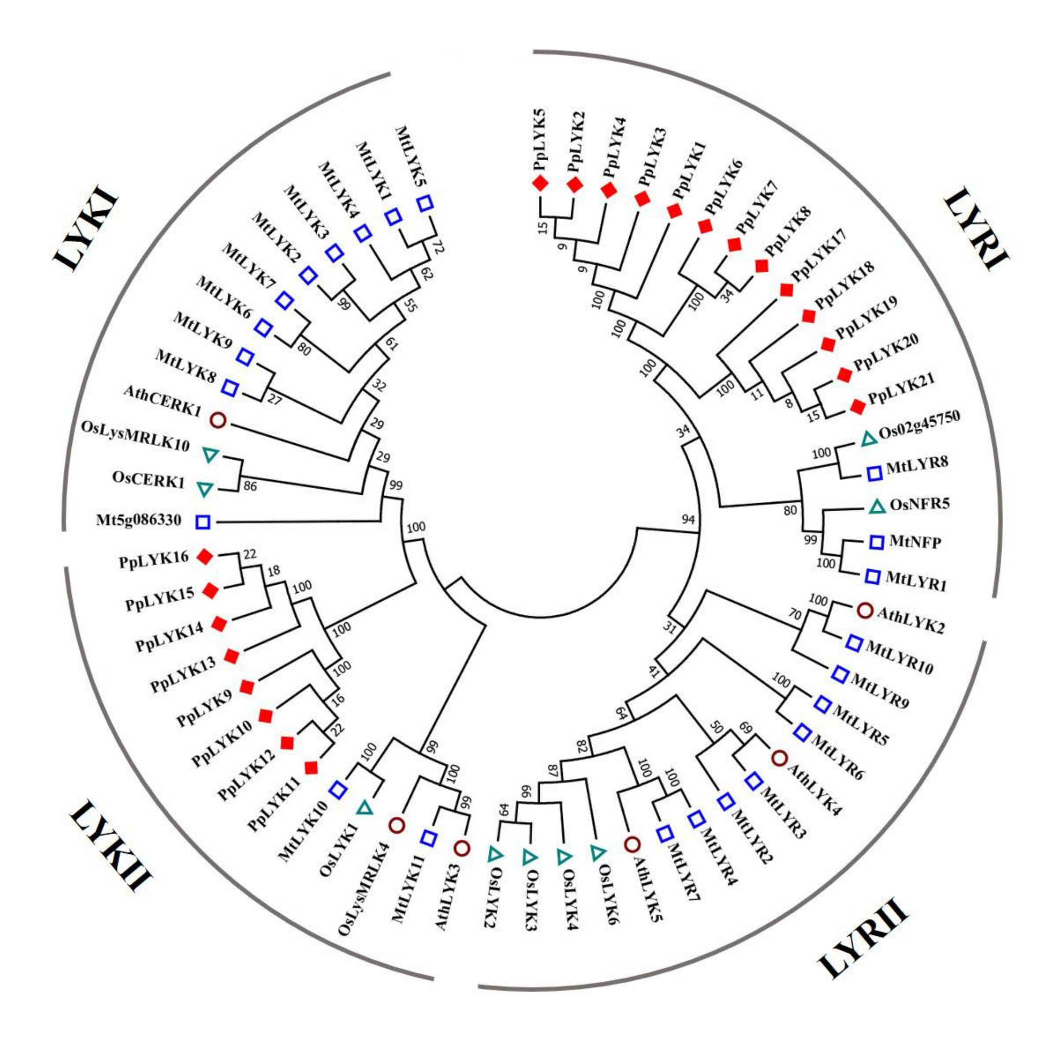
Fig.2 Conservative domain analysis of LYK protein in *Physcomitrella patens*



**2.3 小立碗藓 LYK 蛋白的系统发育分析**

通过 MEGA 对小立碗藓 *LYK* 基因家族成员的蛋白序列进行多序列比对后构建系统发育树（图 1：A），结果显示：其 *LYK* 基因家族的系统发育树有三个非常明显的分支，3 大分支中均发现了独立的进化分支，说明小立碗藓 *LYK* 基因家族成员在漫长的进化历史中产生了不同的进化方式，推测同一大分支延伸出的不同分支成员间功能可能有所差异。

为了进一步将小立碗藓 *LYK* 基因家族成员分类并研究其与其它植物 *LYK* 基因间的进化关系，分别选取 10 个水稻（单子叶模式植物）、5 个拟南芥（双子叶模式植物）和 23 个蒺藜苜蓿（豆科模式植物）中已鉴定出的所有 *LysM-RLK* 基因的氨基酸序列，与小立碗藓 LysM-RLK 共同构建系统进化树，如图 3 所示，分析发现，发现 4 种植物中的 LYK 蛋白存在于 4 个亚族（LYK-I、LYK-II、LYR-I、LYR-II）中，而小立碗藓 LYK 蛋白仅存在于 2 个亚家族：LYK-II 和 LYR-I 中，LYK-II 亚族中有 8 个小立碗藓 LYK 家族成员（*PpLYK9*—*PpLYK16*），另外 13 个成员属于 LYR-I 亚族。LYK-II 亚族中的小立碗藓 LYK 家族成员有两个分支，同源性达 90%以上；LYR-I 亚族中的小立碗藓 LYK 家族成员也存在两个分支，同源性达 90%以上，可能为直系同源基因。



红色菱形框表示小立碗藓 *LYK* 基因，蓝色框表示蒺藜苜蓿 *LYK* 基因，绿色三角形表示水稻 LYK 基因，棕色圆形表示拟南芥 *LYK* 基因。

The red diamond-shaped box represents the *LYK* genes of *Physcomitrella patens*, the blue box represents the *Medicago truncatula LYK* genes, the green triangle represents the *Rice* LYK genes, and the brown circle represents the *Arabidopsis LYK* genes.

图 3 小立碗藓和几种模式植物 LYK 基因系统进化树

Figure.3 Phylogenetic tree of LYK gene of *Physcomitrella patens* and several model plants

**2.4 小立碗藓 *LYK* 基因家族的染色体定位**

通过 Ensembl Plants 数据库获得小立碗藓 LYK 染色体定位信息，将小立碗藓 21 个 LYK 基因定位在 4 条染色体上，如图 4 所示。小立碗藓第 7 条染色体上定位了 9 个基因，为定位 *LYK* 基因数目最多的一条染色体；第 1、9、11 条染色体上分别定位了 3、5、4 个基因，可见小立碗藓 21 个 *LYK* 基因在染色体分布上是相对集中的。图中可明显看到在 4 条染色体上均出现小型基因簇，根据基因簇的定义，基因簇成员在结构和功能上可能具有明显的相似性。其中，4 条染色体上均出现位置接近的基因，如 1 号染色体上基因 *PpLYK6*、*PpLYK7* 和 *PpLYK8*，9 号染色体上的 *PpLYK17*、*PpLYK18*、*PpLYK19*、*PpLYK20* 和 *PpLYK21* 等，可能为串联重复。

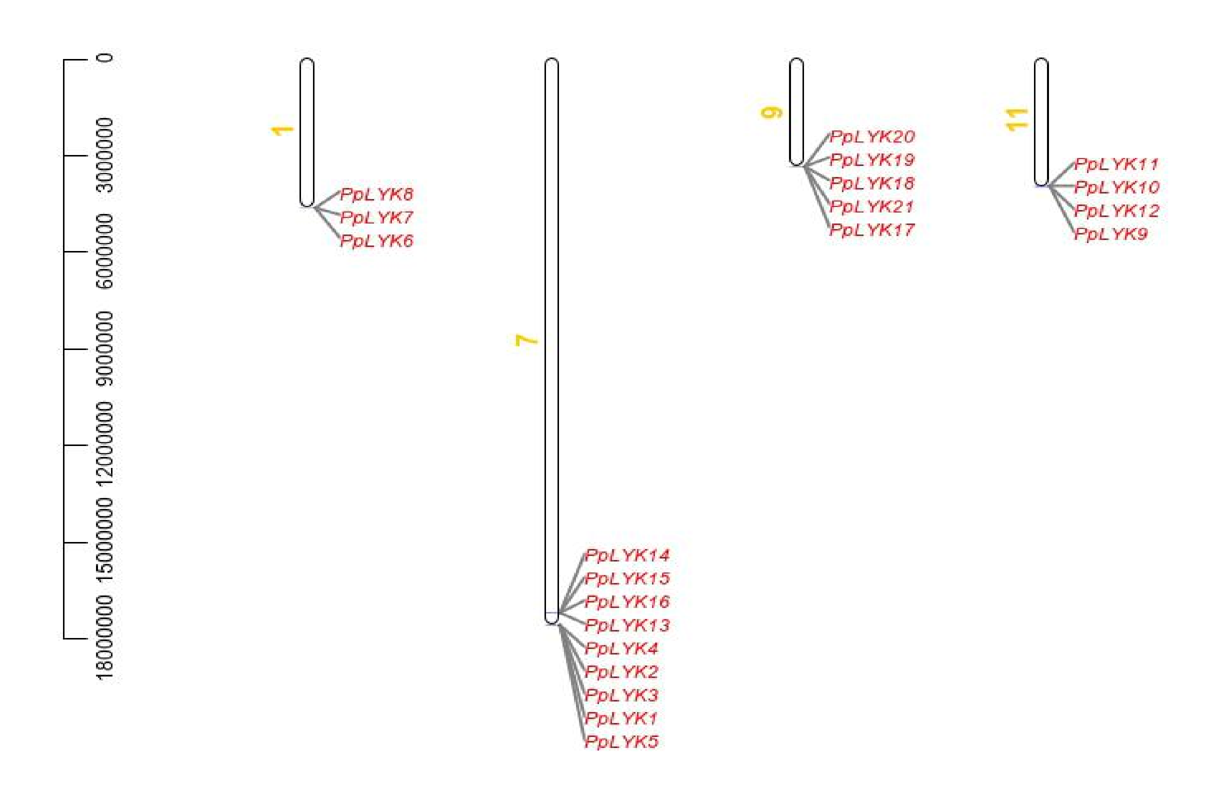



图 4 小立碗藓 *LYK* 基因染色体定位图

Figure. 4 Chromosome localization map of *LYK* gene in *Physcomitrella patens*

## 3 讨论

本文鉴定了小立碗藓基因组中的 *LysM-RLK* 基因，并对其进行生物信息学相关分析。共鉴定出小立碗藓中 21 个 LysM-RLK 家族成员，它们在染色体上的分布相对集中，仅位于 4 条染色体上，且出现 5 组小型基因簇，推测可能为串联重复。每个 LYK 均含有 1~3 个 LysM 保守结构域和 1 个激酶结构域（图 2）,MEME 结果显示同一亚家族各成员具有相同或相似的 domain（图 1：B），说明该亚家族成员可能在小立碗藓整个生命过程中执行相似的功能。

从 LysM-RLK 家族成员数量看，小立碗藓（21 个）多于拟南芥（5 个）和水稻（10 个）中 *LYK* 基因个数，但和蒺藜苜蓿（23 个）同源家族的数量相差不大。研究发现该家族成员在高等植物中普遍存在较多成员，如苹果中有 12 个 LysM-RLK，百脉根（*Lotus japonicus*）中有 17 个 LysM-RLK（周喆等，2014；Lohmann et al., 2010）。小立碗藓是早期登陆的植物代表，其 LysM-RLK 家族成员数量多于拟南芥和水稻，推测该家族成员在小立碗藓中可能存在基因功能的冗余。随着不断进化，一些高等植物中该家族成员数量得到扩展，意味着该家族新功能的引入。

通过系统进化分析，将小立碗藓 *LysM-RLK* 基因分为 2 个亚组（LYK-II 和 LYR-I），它们均与水稻、拟南芥和蒺藜苜蓿同源家族中的部分成员聚为一类，而高等植物中 LysM-RLK 家族分为 4 个亚组， 推断小立碗藓中 LysM-RLK 家族成员的功能相对单一。与进化分析表征相似，同一亚组内，*LYK* 基因的外显子-内含子的排布特征较为接近，同时，此基因结构特征与水稻、拟南芥和蒺藜苜蓿的外显子-内含子分布特征较为

相似（Shiu et al., 2004；Arrighi et al., 2006）。综上所述，同一亚组内基因家族成员具有相近的保守域特征、同源关系和基因结构特征，说明其可能具有相同或相似的功能。

LysM 型类受体蛋白激酶基因在植物生长发育及逆境胁迫中发挥重要功能，已成为当今植物中基因功能研究的热点之一，*LysM-RLK* 基因家族在高等植物中数量较多，功能也具有多样性，但在小立碗藓中功能的研究不够深入，大部分 *LYK* 基因功能尚不明确。本研究以已有数据库资源为基础，在小立碗藓中筛选得到 21 个 *LYK* 基因，并利用生物信息学方法，对其 *LYK* 基因家族成员进行蛋白特征、染色体定位、基因结构及进化关系等分析，将小立碗藓 *LYK* 基因家族进行初步分类，从基因层面显示了小立碗藓的特征，为之后研究 LysM 型类受体蛋白激酶功能提供了理论依据。

**参考文献：**


AFZAL AJ, WOOD AJ, LIGHTFOOT DA, 2008. Plant Receptor-Like Serine Threonine Kinases: Roles in Signaling and Plant Defense[J]. Mol Plant Microbe Interaction, 21(5):507-517.

ARRIGHI JF, BARRE A, AMOR BB, et al., 2006. The Medicago truncatula Lysine Motif-Receptor-Like Kinase Gene Family Includes NFP and New Nodule-Expressed Genes[J]. Plant Physiol, 142(1):265-279.

CHEN CJ, CHEN H, HE YH, et al., 2018. TBtools, a Toolkit for Biologists integrating various biological data handling tools with a user-friendly interface. BioRxiv 289660.

CHEN ZH, 2019. Conservation and functional differentiation of E (z) homologous genes in plants [D]. [陈紫华, 2019. 植物中 E(z)同源基因的保守性和功能分化研究[D].]

KERSEY PJ, 2018. Ensembl Genomes 2018: an integrated omics infrastructure for non-vertebrate species[J]. Nucl Acid Res, (D1):D1.

EDGAR RC, 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput[J]. Nucl Acid Res, (5):5.

FAN YL, 2009. Simple repeated sequence analysis of the genome of *Xanthomonas oryzae pv. Oryzae* [J]. Anhui Agricultural Sciences, No.273 (20): 9387-9390. [樊颖伦, 2009. 水稻白叶枯病菌基因组简单重复序列分析[J]. 安徽农业科学, No.273(20):9387-9390.]

ARRIGHI JF, ANNICK B, BEN AB, et al, 2006. The Medicago truncatula Lysine Motif-Receptor-Like Kinase Gene Family Includes NFP and New Nodule-Expressed Genes[J]. Plant Physiol, 142(1):265-279.

LEHTI-SHIU MD, ZOU C, HANADA K, et al., 2009. Evolutionary History and Stress Regulation of Plant Receptor-Like Kinase/Pelle Genes[J]. Plant Physiol, 150(1):12-26.

LIMPENS E, FRANKEN C, SMIT P, et al., 2003. LysM Domain Receptor Kinases Regulating Rhizobial Nod Factor-Induced Infection[J]. Science, 302(5645):630-633.

LOHMANN GV, SHIMODA Y, NIELSEN MW, et al., 2010. Evolution and Regulation of the Lotus japonicus LysM Receptor Gene Family[J]. Mol Plant Microbe Interaction, 23(4):510-521.

POOLE RL, 2007. The TAIR database[J]. Methods Mol Biol, 406(406):179-212.

RADUTOIU S, MADSEN LH, MADSEN EB, et al., 2003. Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases[J]. Nature, 425(6958):585-592.

RENSING SA, ROMBAUTS S, PEER YVD, et al., 2002. Moss transcriptome and beyond. Trends Plant Sci, 7(12):535-538.

STONE J, 1995. Plant protein kinase families and signal transduction[J]. Plant Physiol, 108(2):451-457.

SHIU S, KARLOWSKI WM, PAN R, TZENG Y, MAYER KFX, LI W, 2004. Comparative analysis of the

Receptpes-Like kinase family in *Arabidopsis* and *rice*[J]. Plant Cell, 16: 1220-1234.

SONG XQ, 2016. Cloning and functional analysis of the super-accumulating SaCAD gene of Jingtiandongtian [D]. [宋希茜, 2016. 超积累型东南景天 SaCAD 基因的克隆及其功能分析[D].]

SAGULENKO P, PULLER V, NEHER RA, 2018. TreeTime: Maximum-likelihood phylodynamic analysis[J]. Virus Evol, 4(1):1.

SHIU SH, KARLOWSKI WM, PAN R, et al., 2004. Comparative Analysis of the Receptor-Like Kinase Family in Arabidopsis and Rice[J]. Plant Cell, 16(5):1220-1234.

TIAN LM, 2016. Identification, expression pattern analysis and functional study of tomato IMPα / β and LYK family genes [D].[田丽梅, 2016. 番茄 IMPα/β和 LYK 家族基因的鉴定、表达模式分析和功能研究[D]. ]

WAN J , ZHANG XC , NEECE D, et al., 2008. A LysM Receptor-Like Kinase Plays a Critical Role in Chitin Signaling and Fungal Resistance in *Arabidopsis*[J]. Plant Cell Online, 20(2):471-481.

WANG L, LI G, WANG J, et al., 2009. Molecular cloning and evolutionary analysis of the GJA1 (connexin43) gene from bats (Chiroptera)[J]. Genet Res, 91(02):101.

ZHANG XC, 2007. Molecular evolution of Lysin motif-type receptor-like kinase in plants[J]. Plant Physiol, 144.

ZHOU Z, ZHANG CX, ZHANG LY, et al., 2014. Bioinformatics and Expression Analysis of the LysM Gene Family in Apple [J]. Chinese Journal of Agricultural Sciences, 000 (013): 2602- 2612. [周喆，张彩霞，张利义，等，2014. 苹果 LysM 基因家族的生物信息学及表达分析%Bioinformatics and Expression Analysis of the LysM Gene Family in Apple[J]. 中国农业科学, 000(013):2602-2612.]